DOCUMENT RESUME

ED 118 703 UD 015 744

AUTHOR Balasubramonian, K.; And Others
TITLE Do Bilingual Education Programs Inhibit English Language Achievement? A Report on an Illinois Experiment.
PUB DATE May 73
NOTE 30p.; Paper presented at the Seventh Annual Convention - Teachers of English to Speakers of Other Languages (San Juan, Puerto Rico, May 9-13, 1973)

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage
DESCRIPTORS Achievement Gains; *Achievement Rating; *Bilingual Education; Bilingual Schools; Bilingual Students; Comparative Analysis; Early Childhood Education; English (Second Language); Minority Group Children; Primary Education; Program Effectiveness; *Program Evaluation; *Second Language Learning; *Spanish Speaking
IDENTIFIERS *Chicago (Illinois)

ABSTRACT

The English language achievement of 213 Spanish speaking students in grades K-3, who receive English as a Second Language (ESL) instruction in the context of a bilingual program, is compared to that of 104 similar students who receive ESL instruction within the traditional school program, in order to probe whether enrollment in a bilingual program retards the learning of English as a second language. ESL achievement, as measured by listening and reading comprehension tests, is compared through ANOVA and partial correlation analysis with a quasi-experimental evaluation design. Anlaysis of data indicates that students learning English in a bilingual program learn just as much English as students learning it through ESL classes within a traditional curriculum. Because the amount of measured English achievement in kindergarten and first grade is found disappointing compared to the substantial achievement measured in second and third grades, two questions are raised - that of the effectiveness of language instruction in K-1 and that of the appropriateness of the techniques used to measure achievement on these levels. Although the study needs replication with other age groups and other languages, the implication derived from the study is that half-day bilingual programs do not inhibit Engllish language achievement in primary-aged children. (Author/AM)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED118703

(Prepublication Draft)

DO BILINGUAL EDUCATION PROGRAMS INHIBIT ENGLISH LANGUAGE ACHIEVEMENT?: A REPORT ON AN ILLINOIS EXPERIMENT*

K. Balasubramonian

Illinois Bilingual Education Service Center

H. Ned Seelye

Office of the Superintendent of Public Instruction

188 West Randolph

Chicago, Illinois

Rafaela Elizondo De Weffer

Chicago State University

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

A B S T R A C T

Children enrolled in an Illinois bilingual program typically are exposed to approximately 25% less English during the school day than their counterparts in traditional school programs. This raises fears among some educators and parents that enrollment in a bilingual program might retard the learning of English as a second language (ESL). To probe this, pre-post test data were collected in three cities during a five month interval in 1972 on the ESL achievement of 213 kindergarten through third grade Spanish-speaking children taught ESL within the context of a half-day bilingual program and 104 similar children receiving ESL instruction as part of the traditional school curriculum. ESL achievement, as measured by listening and reading comprehension tests, is compared through ANCOVA and partial correlation analysis within a quasi-experimental evaluation design. Analyses indicate no statistically significant difference in ESL achievement between the two groups. The implication of these results is that half-day bilingual programs do not inhibit English language achievement in primary-aged children.

UD 015744

BACKGROUND

The most obvious educational need for children with limited fluency in English who live in continental United States is to learn more English. To this end, at least four didactic policies have been tried.

The easiest policy to implement has been simply to absorb the children whose knowledge of English is limited into the traditional curriculum in the hopes that total school-day exposure would lead to assimilation. In schools with more than a handful of limited-English-speaking children this approach has not lowered dropout rates, or raised achievement scores.

To strengthen the total exposure to English policy, administrators identified limiting variables and proposed ways to eliminate them. Numerous schools in Texas and California, for example, took similar steps to avoid what they perceived to be the debilitating variables. The most widespread measures included the prohibition of speaking a language other than English in the classroom, in the corridors, on the playground. Some well-meaning school administrators even went so far as to advise parents to speak broken English at home to their children, rather than fluent Spanish, to use the most common example, (Krear, 1969). Worse, many parents complied. Unfortunately, this approach did not lower dropout rates or substantially increase test scores

---

*Paper presented at the Seventh Annual Convention, Teachers of English to Speakers of Other Languages, May 9-13, 1973, San Juan, Puerto Rico. It benefited from critical readings by Ann Garcelon, Bilingual Education Section, Illinois Office of Public Instruction, and Marilynn B. Brewer, Psychology Department, Loyola University of Chicago.

of English achievement either. It did, on the other hand, inhibit fluency in the child's home language and increase the level of school misunderstanding. All in all, this approach exacted a high toll from children without an accompaning gain in scholastic performance (Krear, 1969). Millions would be spent by Congress through the National Defense Education Act in an effort to regain some of the "foreign" language fluency lost in great part because of a short-sighted school policy to "mainstream" children.

While the Second World War was forcing the U. S. to reevaluate the worth of second-language fluency, anthropological linguists were developing a technology of language description and instruction which would revolutionize the teaching of English as a second language (kavetsky and Morrison, 1960). There subsequently emerged in our school systems a third approach to teaching English to children whose primary language was not English. Special courses were designed which took into vigorous consideration the fact that the critical learning steps for a child who is studying English as a second language differ from those steps involved in increasing the English of children who already possess native fluency. Not surprisingly, this approach (ESL, TESL, TESOL) has been relatively effective.

By the mid 1960's, political pressures from minority ethnic groups throughout the U.S., especially among the Spanish-speaking in the Southwest, were redefining the context within which much ESL instruction would occur. These ethnic minorities pointed to the scandalously high dropout rate of their children, to the misplacement of grossly disproportionate numbers of their children in classes for the retarded, to their children's loss of self-esteem and pride in their first language and heritage.

Largely as a result of insights formed within the non-English-speaking or bilingual communities, six conditions have been identified which contribute to the hapless syndrome in which the limited-English-speaking child finds himself. First, the student does not know enough English to understand the concepts being taught in the classroom so he fails in his classwork. Second, when enough time

is taken out of the school day to offer him intensive instruction in English, he slips behind in the subject-matter concepts of the classes he has missed, so he still fails when he returns to classes in mathematics, science, and social studies. Third, if in the process of learning English he is allowed to forget his first language, he suffers alienation from his home and ethnic identity, with subsequent loss of pride in his heritage and depletion of this country's bilingual resources. Fourth, any success that he experiences while working exclusively with peers from the same ethnic and language background may not transfer when he has to associate with Anglo (black and white) Americans. Fifth, even if Anglo American children and teachers who have or who develop the necessary empathy to work successfully with limited-English-speaking children are incorporated in a special program, children and especially teachers outside of the special program, with whom the bilingual children must eventually relate, often lack the sensitivity required to deal successfully with children from divergent ethnic backgrounds--no matter how well these latter speak English. And sixth, because of different priorities and past experience with incommunicant school systems, parents often place greater value on having their children work than on having them receive a high school diploma.

Bilingual education attempts to intervene in this discouraging syndrome to effect curriculum changes which deal in a positive fashion with the educational needs of children with limited fluency in English. One aspect of this curriculum is the teaching of English as a second language. (Other components of bilingual education include the teaching of subject-matter concepts through the students' first language, the teaching of the students' first language as a continuing concern, and the teaching of cultural heritage.)

The logic used to justify bilingual education is by no means universally accepted. The most common counterargument states that the most pressing need for a person living in the U.S. is not increased self-esteem or subject-matter knowledge, but fluency in English. Many educators and parents fear that extended school exposure to the child's first language will retard his development in English. A child in an Illinois bilingual program typically is exposed to English only 75% of the school day, whereas the child who is taught ESL as part of the traditional school curriculum hears English virtually the whole school day. It is to this question of how English language achievement is affected by bilingual programs that the scope of the present paper is limited.

Two recent studies deserve mention. Valencia (1971) compared the pre and post test achievement during a four month interval of mostly Mexican, Puerto Rican, and Cuban children enrolled in 15 Chicago schools. Using one-way analysis of variance, Valencia concluded that "the English language development of children with limited or no knowledge of English is being enhanced by the Itinerant ESL, ESL kits, and TESL programs" (p.64). However, since the gains in achievement scores were not compared to similar children who did not receive ESL, this study does not in any way diminish the strength of several rival explanations of the data: that over any four-month period of living in the U.S. and being exposed to English, students will demonstrate some growth in English; or that the increased maturity of the child will explain his ability to score higher at posttest; or that the pretest was reactive and accounts for the posttest gains; and so forth (see Campbell and Stanley, 1963). Zirkel, (1972) conducted a Connecticut study in which he compared Puerto Rican children, first grade through third grade, who were taught ESL within a bilingual education program and children

who received ESL as part of the traditional curriculum. The children were matched according to socio-economic status, sex, age, and exposure to English. Most of the results showed no difference in ESL achievement among the groups.

The advent of bilingual programs in mainland U.S. public schools has been a recent event, beginning in 1968 with ESEA Title VII federal funding. The principal source of funding in Illinois for bilingual programs is state revenues. Last academic year (1971-72), there were 10 state-funded bilingual programs in Chicago and 11 others in school districts outside of Chicago. All of these "downstate" programs were limited to kindergarten through third grade. This year (1972-73), there are 49 state-funded programs in Illinois, and double this number is anticipated next year. The question this study probes is raised with increasing frequency as schools consider bilingual programs as an alternative educational model for children with limited fluency in English.

EXPERIMENTAL DESIGNS AND SAMPLE CHARACTERISTICS

A pretest-posttest nonequivalent control group design was used in this study. Included in the sample were students enrolled in ESL classes in 15 schools of three "downstate" Illinois school districts with large total student enrollment as well as large Spanish-speaking enrollments.

All 317 children included in the sample received instruction in English language arts as part of the regular school curriculum, and all 317 children received in addition specialized ESL instruction for 30-40 minutes daily. The bilingual program students received the regular English instruction in the half-day they attended the traditional curriculum, and they received ESL as part of the bilingual program. The bilingual program children heard, on the average, 25% less English during the school day than did the children who were taught ESL within the traditional curriculum. It is this difference--the

Spanish-speaking context of approximately a quarter of the school day for children enrolled in bilingual programs--which we posit as the only variable substantially differentiating the ESL/bilingual program children from the ESL/traditional program children. This statement, of course, needs some defense.

Comparability of groups.

In an attempt to establish the initial equivalency of the two comparison conditions, the following variables were examined vis a vis the twelve standard threats to design validity: the selection process used to designate a school or classroom to one of the two conditions, the pretest equivalence of the two groups, differential attrition in the two groups, economic level of the schools' studentbody, staff teaching experience, staff ethnicity, Spanish fluency of ESL staff, district size, city population, percent of Spanish-speaking students in districts and sample schools, number of schools in each district, and the ethnic composition of Spanish-speaking students.

Selection. The availability of comparison children (ESL within traditional curriculum), in addition to bilingual program children, within each of the three districts was the result of there being more children in these districts in need of a bilingual program than funds to establish the programs. In each district, the number of students to be included in the bilingual program was pre-determined through proposal negotiation between the district and the state office personnel. Accordingly, the schools which had comparatively higher percentages of Spanish-speaking children were given priority for inclusion in the bilingual program. However, the restriction in numbers imposed by limited funds left some schools or classrooms of comparable Spanish-speaking proportions out of the program. These students also received ESL instruction and were considered comparison students for evaluation purposes. It must be noted that

assignment to the two groups was not random.

There were seven ESL/bilingual schools and eight ESL/traditional schools. All students enrolled in either of these two conditions who were present for both pretesting and posttesting were included in this study.

Pretest equivalence. While pretest scores were not used as a criterion for membership in either group--all students present for both pre and post testing were included in the study--a test of significance was run on the correlation between pretest scores and group membership. There was no statistically significant correlation between pretest scores and the two groups, with the exception of the first grade data which was significant at the .05 level (see Tables I and II).

Attrition. The year began with 422 (289 ESL/bilingual and 133 ESL/traditional) students, but by year's end attrition had reduced the number to 317 (213 ESL/bilingual and 104 ESL/traditional curriculum). The attrition rate in the ESL/bilingual group was 26%, and 21% in the ESL/traditional group.

Economic index of schools. Title I eligibility (1) was taken as an indication of the poverty level of the school community. Of the fifteen schools involved in the bilingual programs, all are Title I eligible schools. Of the ESL/traditional curriculum schools, seven of the eight are Title I eligible.

Staff experience. Students in both groups generally meet with three adults over the course of the day: regular homeroom teacher, ESL teacher (either within a bilingual program or within a traditional curriculum), and a teacher aide. This study groups together teachers and teachers aides. The average teaching experience for the homeroom teachers of both groups was 14 years. The average experience of the ESL/traditional staff was seven years teaching (two of which were in ESL), while the average teaching experience

of the ESL/bilingual staff was three years (one of which was in ESL). The ESL/traditional staff had, on the average, somewhat over double the teaching experience of the ESL/bilingual staff.

Staff ethnicity. Of those teachers and aides working in the ESL/traditional curriculum, six were of English-speaking background and two were of Spanish-speaking background. The ESL/bilingual staff included two from Anglo backgrounds and eight from Hispanic backgrounds. All the homeroom teachers were Anglos.

Fluency in Spanish. None of the homeroom teachers spoke Spanish. All the staff teaching in the bilingual program were fluent in Spanish. Of the ESL/traditional curriculum staff, four were fluent in Spanish and four were not. A separate analysis of the data was made to determine whether there was a significant difference in the achievement of those ESL/traditional curriculum students who had ESL teachers fluent in Spanish was

Data on the size of the three districts and the cities' population, their concentration of Spanish-speaking students, the number of schools in each district, and the ethic composition of the districts' Spanish-speaking students is contained in Table II. Two differences between the two groups is evident. The districtwide percentage of Spanish-speaking students varies widely. On the other hand, the means of the 15 schools in our sample are completely comparable, although the standard deviations do vary. This indicates that there is more variance 'within district scores' than 'between district scores'. The other difference between the two groups is in the ethnic national makeup of the Spanish-speaking students (Mexican, Puerto Rican, Cuban). This fact has not seemed to produce any observable effect on test scores, and the generalizeability of the results is consequently enhanced.

Instrumentation. The inadequacy of standardized tests for measuring achievement of minority children is widely appreciated (Wrightstone, n.d.; Fitzgibbon, n.d.; Navassy, 1972; Zirkel, 1972; Wilson, 1973). All of the shortcomings of these instruments are exacerbated when the testee's first language is not the language of the test. To gain a measure of the tests' homogeneity with Illinois' multi-ethnic Spanish-speaking children, Kuder-Richardson 20 reliability was computed for each of the instruments used in this study. Their reliability, calculated on data from several hundred K-3 Latin children besides those included in the present study, ranged from .54 to .81. Table III presents this data in some detail.

Since reliability is partly a function of the number of test items, those instruments where only half the items were administered (see below) did not achieve as high a reliability as one usually expects from commercial tests. Nonetheless, test reliabilities were sufficiently high to have reasonable confidence in their results. A detailed replication study of instrumentation effects of a dozen different tests, based on a sample of 5,000 Spanish-speaking children in Illinois, is currently being conducted by the authors.

Data collection. Pre and post tests of English language achievement (along with tests of Spanish language, mathematics, and self-concept, which will be reported in a future paper) were given to students in both ESL/bilingual and ESL/traditional groups in January and May of 1972. These dates correspond to the tenure of the first semester of implementation of a bilingual program in these districts. All testing was accomplished within a two-week period by bilingual teachers and/or aides who were not the students' regular teachers. For purposes beyond the scope of this study, most tests were administered bilingually. (That is, odd items were given in English and even

items in Spanish, and vice versa). Only the English language portion of the data is considered for this study.

Data analysis. Analyses of data from pre and post measures to ascertain gain have always been a matter of controversy. Unreliability of tests, statistical regression to the group mean phenomena, and correlation between gain scores and pretest scores are the three major problems in the analysis of such data.

Analysis of covariance (ANCOVA) and partial correlation analysis are two of the techniques suggested for overcoming these difficulties. However, it should be noted that ANCOVA has been criticized for underadjusting for regression effects (Campbell and Erlebacher, 1970a). Users of ANCOVA should recognize the procedure of adjustment available (Lord, 1968) and some cautions that are to be maintained (Campbell and Erlebacher, 1970b). Partial correlation analysis is criticized for its inability to correct for differences among groups arising from non-randomization (Lord, 1963; Brewer, Campbell Crano, 1970
Some adjustments in covariance appropriate to different evaluation designs have been suggested (Porter, 1973). Research is underway to taxonomize methods to measure change in terms of practical applicability (Balasubramonian).

The present experiment is a situation where the covariate is fallible due to the imperfect reliability of the pretest measure, and there is a possibility of a systematic difference on the covariate due to non-random assignment. As a procedure of cross-validation, both ANCOVA and partial correlation analysis were performed, and Lord's technique of adjustment was applied to the second and third grade data, using verbal and non-verbal intelligence scores as additional covariates.

DESIGN VALIDITY

Campbell and Stanley (1963) identify eight potential threats to internal validity and four threats to external validity in any experiment. Internal validity asks the question of whether the treatment made a difference in the experiment; external validity asks how generalizable are the experiment's results. A brief review of the information presented in the preceding section, in terms of the potential threats to validity, will facilitate an evaluation of whether the results of this experiment can be interpretable.

Internal validity.

Since both groups shared the same school districts during the same time span, any artifact caused by history (events occuring between the two test periods) could be expected to be shared by both groups. The possible confound which might be speculated due to the differential experience and ethnicity of the teaching staff is mitigated by the fact that each student receives instruction from at least three teachers.

Maturation (growing older) and testing (the effect of taking a test on test scores) are likewise controlled because they would not operate differentially between the two groups.

Instrumentation threats (changes in the calibration of a test or tester) are minimized by having used standardized instruments whose reliability has been calculated, and by using a variety of test administrators for both groups. Testing usually took two days to accomplish and the administrative procedures were the same for both groups. Test-retest effects were not determined; a separate study in progress will do this.

Regression artifacts (normal regression toward the group mean upon retesting) are ruled out for several reasons. First, students were not selected on the basis of pretest scores. Whatever statistical regression does occur should be present equally in both groups. Further, the pretest equivalence

supports the absence of differential regression effects. The dual use of ANCOVA and partial regression analysis affords yet another adjustment for possible regression effects.

Since randomization did not occur, selection is a serious potential threat to internal validity. To be compared, the students in both groups should be the product of the same selection process. In this study, selection was a function of attendance by classroom and/or building. The cutoff point that determined which schools would implement bilingual programs was to a considerable extent the result of the limited availability of funds. The pretest scores of the two groups did not differ significantly. The teaching experience of the staff favored somewhat the ESL/traditional classroom setting. School size, concentration of Spanish-speaking students, numbers of staff involvement in ESL, geographical location of schools, Spanish fluency among the ESL staff, did not substantially differ between the groups. We feel the plausibility of selection artifacts to be minimal.

Experimental mortality (differential attrition from the two groups) appears to be comparable in terms of rate of attrition. This does not rule out bias in the type of student who left the two groups, but we are unaware of any systematic bias that would not be true across both ESL/bilingual and ESL/traditional schools.

The final threat to internal validity, interaction effects among any of the seven other threats, appears extremely unlikely given the preceding information.

External validity.

The threat of interaction of testing and treatment is not controlled by this design. It is, however, somewhat contained by the fact that most of the students' immediate school environment is exposed to the same condition and testing is an accepted school activity. At any rate, the generalizability of

the results is not meant to extend beyond children learning ESL in the settings in which they were tested (i.e., bilingual programs or traditional programs). A design variation in which random schools within both groups would have the pretest omitted would offer a control for this confound.

The threat of interaction of selection and treatment is partly diffused by our willingness to limit the applicability of the results to students who would have been eligible for both ESL and bilingual programs on the primary school level. This eligibility is based largely on language need. There is an apparent danger in extending the interpretability of the results beyond the type of child who might have attended one of these two Illinois programs. On the other hand, the multi-ethnic nature of Illinois' Spanish-speaking population is ideally suited to generalizing across Spanish-speaking groups. There is the possibility that the schools which successfully requested for bilingual programs had administrators who were more open to the educational needs of limited-English-speaking children. Our experience has been mixed in this regard, and we do not think that there is a strong systematic bias favoring bilingual program children for this reason.

The third threat to external validity Campbell and Stanley call reactive arrangements. These would be especially serious in artificial experimental conditions. This reactive effect might tend to favor a child who envisioned himself in an "exotic" or especially prestigious program. It is doubtful that this was the case with the two groups reported in this study. At any rate, the results are not meant to be generalized beyond the described conditions, however reactive.

The final threat to the generalizability of an experiment's results, multiple treatment interference, raises the question of what would happen to the English language achievement of either group once the continuous

nature of the ESL is substantially altered in structure, as in the case of an alternate-days, rather than daily, program. The present study does not generalize beyond a daily treatment effect. The carry-voer power of ESL instruction has not been probed here.

The appropriateness of this nonequivalent control group design for the circumstances in which it was employed is judged by the authors to be capable of providing eminently interpretable data. The few instances of possible bias cancelled each other out by their contrastive direction. The two groups seem remarkably similar.

RESULTS

The question raised in the title of this paper is whether bilingual education programs inhibit English language achievement.

To probe this question, English language achievement of 213 Spanish-speaking students in grades K-3 who received ESL instruction in the context of a bilingual program was compared to that of 104 similar students who received ESL instruction within the traditional school program. Tables 4, 5 and 6 present the ANCOVA results of the three grade levels* Table 7 presents the results of partial correlation analysis for the three grades.

An examination of the tables indicates that in both types of analyses, regardless of grade level, the students studying ESL performed equally whether they were learning it within the context of a bilingual program or a traditional program. In ANCOVA, none of the F statistics was significant. Similarly, in partial correlation analysis, none of the correlations was significant. In the case of second and third grade data, even when intelligence (as measured by the Inter-American Series Test of General Intelligence Level II) and pretest performance were partialled out simultaneously, the correlations were not significant.

The results all indicate that students learning English in a bilingual program learn just as much English as students learning it through ESL classes within a traditional curriculum.

IMPLICATIONS FOR FURTHER STUDY

The educational significance of the amount of measured English achievement in kindergarten and first grade is disappointing, while the achievement measured in second and third grades appears substantial. This raises at least two questions: the effectiveness of language instruction in K-1, and the appropriateness of the techniques we used to measure English achievement on these levels. The answer more likely lies with the measurement techniques. It was especially difficult to test children of this tender age.

It is frequently argued that language ability, especially on the K-3 level, is principally a matter of oral skills; and oral ability should be measured through tests which evoke oral responses and are in themselves oral. There is no philosophical quarrel with this position. The economics of testing force the investigator to choose between a large sample tested on listening and reading comprehension items, or a small sample tested through taped oral responses. Prompted by expediency, and encouraged by the high correlations between the briefer written form of the MLA Proficiency Test and the longer form with an oral component, our oral tests opted for a large number of children answering questions of known reliability. A controlled study using direct measures of oral language, would be a distinct contribution to the literature.

It may be argued that the reason for the "no significant difference between comparison groups" is that the posttest results were the product of some pre-bilingual program learning which was still exerting its effect. Since the tested

interval covered only four months time, this rival explanation of the data has mild plausibility. A longitudinal study should elucidate this issue, and one is currently being conducted by the authors.

Can this study's findings be extended to other language groups and to children of other ages? We would hypothesize a yes to both questions, but there is a need to replicate this study's findings in other language and age groups.

Would the results have been the same had we compared students studying in a traditional English language arts program (not ESL) to similar students studying ESL within either a traditional academic setting or within a bilingual program? A controlled study should elucidate this, too.

There is always difficulty "proving" a null hypothesis. It should be noted, however, that the use of covariates substantially increased the power of the analysis to detect small differences had they been there. Nonetheless, in spite of the questions we have raised, the overwhelming implication of this study is that a student learns every bit as much English in a bilingual program as he does in a traditional all-English school program. When one adds to this the other educational objectives of a bilingual program, there is ample reason to regard bilingual programs with spirited interest.

REFERENCES

Balasubramonian, K. Simulative and Empirical Analysis of Strategies for Measuring Change and Change Related Characteristics, Doctoral dissertation in progress at Michigan State University.

Brewer, M.B., Campbell, D. T., Crano, W. D., Testing a single-factor model as an alternative to the measure of partial correlations in hypothesis-testing research. Sociometry 33 (1), 1970, p.1-11.

Campbell, D.T. and Erlebacher, A., How Regression artifacts in Quasi Experimental evaluations can mistakenly make compensatory education look harm ful., in J. Hellmuth (ed), Compensatory Education: A national debate, Vol. 3 Disadvantaged Child, (New York: 1970a) Brunner/Mazel, p. 185-210

Campbell, D. T. and Erlebacher, A. (1970b) Note on the comments by Cronbach and Porter on Common Factor Covariate Adjustment mentioned in Campbell and Erlebacker (1970a).

Dubois, P.H., Mayo, C.D.(ed) Research Strategies for Evaluation Training, AERA Monograph on Curriculum and Evaluation - 4. (Chicago:1970) Rand McNally and Co., Chapters 2 and 3.

Fitzgibbon, T.J. The use of Standarized instruments with urban and minority group pupils, n.d., Test Department, Harcourt Brace Jovanovich Inc.

Havassy, Barbara A critical review of the New Inter-American Series, in Bilingual Testing and Assessment, Proceedings of the BABEL workshop, 1972, p.43-105.

Krear, Serfina E. The role of the mother tongue at home and at school in the development of bilingualism. English Language Teaching 24 : 2-4, 1969.

Lord, F.M. Elementary models for measuring change. in C. W. Harris (ed) Problems in Measuring Change, (Madison:1963) University of Wisconsin Press, 1963

Lord,F.M. Statistical adjustments when comparing pre-existing groups. Psychological Bulletin, Nov. 1969, 72 5, p.336-337

Manuel, Herschel T. Technical Report: Tests of General Ability and Tests of Reading - Inter-American Series forms CE,DE,CEs,& DEs. (Austin:1967), Guidance Testing Associates, p.171-21

Moss, Margaret H. Examiners' Manual: Test of Basic Experiences - Level K (Monterrey:1971), CTB/McGraw Hill, p.46

Porter,A.C. Analysis strategies for some common evaluation paradigms. Paper presented at the 1973 AERA convention

Valencia,A.A. An assessment and evaluation study of the Chicago Public Schools ESL/Bilingual programs for non-English speaking children (New Mexico:1971), Southwestern Cooperative Educational Lab. Inc. pp.76

Wilson,J.W. Standardized tests very often measure the wrong things. Mathematics Teacher, April 1973, p.295, 367-370

Wrightstone,J.W.; Hogan,T.P.; and Abbot,M.M. Accountability and associated measurement problems, n.d. Test Department, Harcourt Brace Jovanovich Inc.

Zirkel,P.A. An evaluation of the effectiveness of selected experimental bilingual education programs in Connecticut - Doctoral dissertation, University of Hartford, 1972, pp.164

Zirkel,P.A. The standardized testing of Spanish-speaking children, Urban Review, 1972

TABLES

TABLE 1

Correlations between student measures and group membership

	PRE ENG.	POST ENG.	GROUP	
PRE ENG.	-			
POST ENG.	.25	-		KINDERGARTEN
GROUP	.05	.16	-	

DF = 99

	PRE ENG.	POST ENG.	GROUP	
PRE ENG.	-			
POST ENG.	.27	-		FIRST GRADE
GROUP	.21*	.17	-	

DF = 104

	PRE ENG.	POST ENG.	GROUP	PRE V.I.	POST N.V.I.	
PRE ENG.	-					
POST ENG.	.79	-				
GROUP	.04	.13	-			SECOND AND THIRD GRADES
PRE V.I.	.33	.27	.14	-		
PRE N.V.I.	.34	.30	.08	.40	-	

DF = 108

* Significant at .05 level

TABLE 2

Selected data on district comparability

DISTRICT	POPULATION	# OF PUPILS IN ELEM. SCHOOLS	DISTRICT-WIDE % OF SPANISH-SPEAKING PUPILS	# OF SCHOOLS* INCLUDED IN STUDY	MEAN % OF SPANISH-SPEAKING PUPILS	S	#OF SCHOOLS BEYOND +/- 1 SD	ETHNIC COMPOSITION
1	80,378	12,557	12	5	21	6	1	100% Mex.
2	65,000	10,929	6	5	24	15	2	33% Mex. 33% P.R. 33% O.L.A.
3	55,691	7,538	2	5	23	9	1	50% Mex. 50% P.R.

* All but one school are Title I eligible.

TABLE 3

Tests and reliability data

TEST NAME	SUBTEST (IF ANY)	PUBLISHER REPORTED				ILLINOIS DATA			
		n	TYPE OF RELIABILITY	r	SE	n	TYPE OF RELIABILITY	r	SE
TEST OF BASIC EXPERIENCES-K	LANGUAGE	2,615	KR20	.82	2.10	89	KR20	.55	1.66
TEST OF BASIC EXPERIENCES-L	LANGUAGE	1,701	KR20	.78	2.01	80	KR20	.55	1.68
TEST OF READING (INTER-AMERICAN, II DE)	VOCABULARY	207	Adjacent forms	.49	5.40 SD	125	KR20	.79	1.77
	SPEED	207	Adjacent forms	.72	5.5 SD	125	KR20	.54	1.49
	LEVEL	207	Adjacent forms	.73	6.1 SD	125	KR20	.81	1.66
TEST OF GENERAL INTELLIGENCE -HABILIDAD GENERAL	VERBAL - NUMERICAL	207	Parallel	.70	5.3 SD	122	KR20	.55	1.51
	NON-VERBAL	207	Parallel	.64	5.2 SD	122	KR20	.64	1.47

TABLE 4

Analysis of covariance table: Kindergarten data

SOURCE	DF	YY	SUM-SQUARES (DUE)	SUM-SQUARES (ABOUT)	DF	MEAN-SQUARE
TREATMENT (BETWEEN)	1	17.3828				
ERROR (WITHIN)	99	660.5781	40.3330	620.2449	98	6.3290
TREATMENT + ERROR (TOTAL)	100	677.9609	42.6554	635.2654	99	
DIFFERENCE FOR TESTING ADJUSTED TREATMENT MEANS....				15.0205	1	15.0205

NULL HYPOTHESIS. NO DIFFERENCE AMONG TREATMENTS AFTER ADJUSTING WITH COVARIATES.

F(1, 98)= 2.373

TABLE 5

Analysis of covariance table: First grade data

SOURCE	DF	YY	SUM-SQUARES (DUE)	SUM-SQUARES (ABOUT)	DF	MEAN-SQUARE
TREATMENT (BETWEEN)	1	23.1953				
ERROR (WITHIN)	104	816.6641	46.7199	769.9441	103	7.4752
TREATMENT + ERROR (TOTAL)	105	839.8594	59.2741	780.5852	104	
DIFFERENCE FOR TESTING ADJUSTED TREATMENT MEANS....				10.6411	1	10.6411

NULL HYPOTHESIS. NO DIFFERENCE AMONG TREATMENTS AFTER ADJUSTING WITH COVARIATES.

F(1, 103)= 1.424

TABLE 6

Analysis of covariance table: Second and third grades data

SOURCE	DF	YY	SUM-SQUARES (DUE)	SUM-SQUARES (ABOUT)	DF	MEAN-SQUARE
TREATMENT (BETWEEN)	1	858.7500				
ERROR (WITHIN)	108	51249.8750	32440.0781	18809.7969	105	179.1409
TREATMENT + ERROR (TOTAL)	109	52108.6250	32859.5820	19249.0430	106	
DIFFERENCE FOR TESTING ADJUSTED TREATMENT MEANS....				439.2461	1	439.2461

NULL HYPOTHESIS. NO DIFFERENCE AMONG TREATMENTS AFTER ADJUSTING WITH COVARIATES.

F(1, 105)= 2.452

TABLE 7

Results of partial correlation analysis*

STATISTICS	K	1	2 - 3				
	PRE ENG.	PRE ENG.	PRE ENG.	PRE V.I.	PRE N.V.I.	PRE ENG. & PRE V.I.	PRE ENG. & PRE N.V.I.
Partial r.	.153	.116	.154	.095	.150	.152	.093
DF	98	103	107	107	107	106	106
P	.127	.236	.111	.325	.121	.116	.341

* The correlations in the table are between posttest and group membership. The variables listed in the table are partialled out.

TABLE 8

Summary data

GRADE	GROUP	PRE ENGLISH MEAN	POST ENGLISH MEAN	PRE VERBAL INTELLI-GENCE MEAN	PRE NON-VERBAL INTELLI-GENCE MEAN	POST ENGLISH MEAN ADJUSTED FOR COVARIATES	STANDARD ERROR	n
K	ESL/BILINGUAL	7.12	8.22	-	-	8.21	.33	101
	ESL/TRADITIONAL	6.32	7.16	-	-	7.22	.37	
1	ESL/BILINGUAL	8.87	10.29	-	-	10.17	.35	106
	ESL/TRADITIONAL	7.80	9.34	-	-	9.51	.42	
2-3	ESL/BILINGUAL	29.39	54.22	25.57	49.93	53.62	1.62	110
	ESL/TRADITIONAL	26.51	48.44	24.10	42.24	49.44	2.10	

TABLE 9

Composite summary data

GRADE	PRE ENGLISH		POST ENGLISH		PRE VERBAL INTELLIGENCE		PRE NON-VERBAL INTELLIGENCE		n
	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.	
K	6.97	6.94	8.01	2.60	-	-	-	-	101
1	Composite not calculated due to significant initial differences between the two groups								
2-3	43.30	18.53	52.06	21.86	28.32	10.09	25.02	9.32	110